

ВВЕДЕНИЕ

Система «STATISTICA», разработанная компанией StatSoft, является одной из наиболее популярных статистических программ для поиска закономерностей, прогнозирования, классификации, визуализации данных. Может применяться в экономике, промышленности, медицине, научных исследованиях и других сферах человеческой деятельности. Клиентами StatSoft являются крупнейшие компании с мировым именем.

В системе существует возможность проводить классические и новейшие методы проведения анализа данных: кластерный, факторный, корреляционный, дисперсионный анализ, линейную и нелинейную регрессии, нейронные сети и др. Визуализация исходных, промежуточных, выходных данных может быть осуществлена выбором из большого числа различных графиков, пиктографиков и диаграмм. Применение программы «STATISTICA» позволяет эффективно решать сложные проблемы и осуществлять аналитическую поддержку принятия решений.

Система «STATISTICA» получает наивысшие оценки во всех сравнительных обзорах статистического программного обеспечения.

СПРАВОЧНЫЕ ДАННЫЕ

1. Краткие сведения из теории статистики

Статистика позволяет компактно описать данные, понять их структуру, провести классификацию, увидеть закономерности в хаосе случайных явлений.

Переменная (variable) – это то, что можно измерять, контролировать или чем можно манипулировать в исследованиях. Общее представление о значениях переменной дают описательные статистики:

- минимум;
- максимум;
- среднее (сумма значений переменной, поделенная на число значений);

$$\bar{x} = \frac{1}{n} \sum_{j=1}^n x_i, \quad \sum_{i=1}^n (\bar{x} - x_i) = 0 ;$$

- дисперсия (изменяется от нуля до бесконечности, нулевое значение показывает, что переменные постоянны, изменений нет);

$$\sigma^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 ;$$

- стандартное отклонение (это корень квадратный из дисперсии; чем выше дисперсия и стандартное отклонение, тем сильнее разбросаны значения переменной относительно среднего).

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} .$$

Два вида случайных величин:

- независимая – x (фактор);
- зависимая – y (результативный признак).

Для определения аналитического выражения связи между независимой и зависимой случайными величинами используется регрессионный анализ. Форма связи – это уравнение регрессии.

Виды регрессии:

- парная – исследуются две величины: фактор и результативный признак, $y = A + Bx$;
- множественная – исследуются несколько факторов и результативный признак, $y = A_0 + A_1x_1 + A_2x_2 + \dots + A_nx_n$.

Для составления прогнозов может быть использовано уравнение регрессии, в котором определяются коэффициенты, называемые параметрами регрессии. Построенная линия уравнения регрессии показывает тенденцию изменения данных и называется линией тренда. Для создания линии тренда используются следующие виды аппроксимации:

- линейная, $\hat{y} = Ax + B$, где A – тангенс угла наклона прямой, B – точка пересечения с осью ординат;
- логарифмическая, $\hat{y} = C \ln x + B$, где C, B – константы;
- полиномиальная, $\hat{y} = C_6x^6 + \dots + C_1x + B$, где $C_6 \dots C_1, B$ – константы;
- степенная, $\hat{y} = Cx^B$, где C, B – константы;
- экспоненциальная, $\hat{y} = Ce^{Bx}$, где C, B – константы.

Достоверность аппроксимации определяется коэффициентом корреляции R , характеризующим связь между двумя множествами в уравнении линейной регрессии.

Если R лежит в диапазоне от 0,9 до 1, то можно применить линейное уравнение регрессии. Если R близко к минус единице, то между наблюдаемыми зависимостями существует обратная зависимость. При других видах аппроксимации используется индекс корреляции.

При R , равном или меньше, 0,3 присутствует слабая линейная связь; при R , равном 0,3–0,5, – умеренная линейная связь; при R , равном 0,5–0,7, –

средняя или заметная линейная связь; при R , равном или больше 0,7, – сильная или высокая линейная связь; при R , равном или больше 0,9, – очень сильная или весьма высокая линейная связь. При R , равном единице, – полная функциональная зависимость, все точки на прямой.

Для определения степени влияния факторов на результативный признак используется дисперсия следующих видов:

- общая дисперсия, показывающая степень влияния основных и остаточных факторов:

$$\sigma^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n},$$

где y_i – значение результативного признака,

\bar{y} – среднее значение результативного признака,

n – число наблюдений (значений результативного признака);

- факторная дисперсия, показывающая степень влияния основных факторов:

$$\sigma_{\phi}^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{n},$$

где \hat{y}_i – теоретическое значение результативного признака;

- остаточная дисперсия, показывающая степень влияния остаточных факторов, остаток – это разница между реальными значениями и теоретическими прогнозируемыми:

$$\sigma_o^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - (m + 1)},$$

где m – число факторов.

Если существует корреляционная связь, то выполняется соотношение:

$$\sigma^2 = \sigma_{\phi}^2 + \sigma_o^2; \quad \sigma_{\phi}^2 < \sigma^2.$$

При анализе с помощью множественной регрессии используется множественный коэффициент детерминации R^2 , называемый также квадратом коэффициента множественной корреляции R и определяющий долю вариации результативного признака, обусловленную изменением факторов:

$$R^2 = \frac{\sigma_{\Phi}^2}{\sigma^2} .$$

При R^2 больше 0,7 (критерий Фишера) считается, что вариация обусловлена влиянием факторов.

Кроме того, для оценки результата аппроксимации используется средняя ошибка аппроксимации:

$$\bar{\varepsilon} = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \cdot 100\% .$$

Если некоторые параметры регрессии малы по сравнению со стандартной ошибкой, то их можно исключить из анализа.

Для определения значимости результата используется p -уровень, характеризующий вероятность ошибки. Обычно принимается p -уровень равным 0,05 или меньше, т.е. ошибка должна быть не более 5%.

2. Биржа

Биржи публикуют информацию о совершаемых сделках на различные виды активов. В течение биржевой сессии цена изменяется следующим образом:

- цена открытия (Open Price) – цена, по которой совершается первая сделка;
- закрытия (Close Price) – цена, по которой совершается последняя сделка;
- высшая цена (High Price) – цена, зафиксированная в течение сессии;
- низшая цена (Low Price) – цена, зафиксированная в течение сессии.

Эта информация имеет различное представление для разных бирж, но, как правило, цены сообщаются в таком порядке: Open, High, Low, Close.

Цены Open, High, Low, Close на биржевом графике изображаются с помощью Candelsticks (японские свечи). Для каждой даты рисуется фигура (свеча). Цены открытия и закрытия представляются прямоугольником. Если цена открытия ниже цены закрытия, то прямоугольник белого цвета, нижняя сторона соответствует цене открытия, верхняя сторона – цене закрытия. Если цена открытия выше цены закрытия, то прямоугольник черного цвета, нижняя сторона соответствует цене закрытия, верхняя сторона – цене открытия. Этот прямоугольник пересекает вертикальный отрезок. Верхняя точка отрезка соответствует высшей цене, а нижняя – низшей (рис. 1).

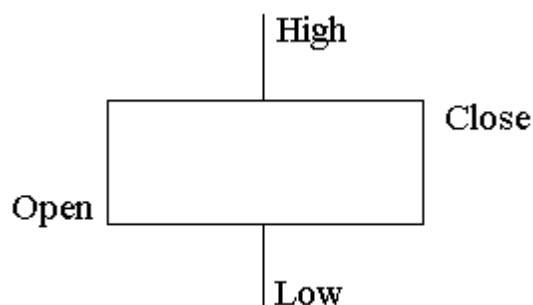


Рис. 1. Candelstick

Фьючерс – это договор купли-продажи базового актива с исполнением обязательств в установленную дату в будущем. Цены на фьючерсные контракты представлены в долларах США, причем используются тридцать вторые доли доллара. Так, цена фьючерса 93-14 означает $93 \frac{14}{32}$, т.е. 93,4375. Чтобы получить цену контракта, эту величину надо умножить на 1000.

3. Пользовательский формат в Excel

Для создания пользовательского формата в Excel применяется команда «Ячейки» из меню «Формат». На вкладке «Число» из списка форматов

выбрать строку «(все форматы)». В строке «Тип» ввести шаблон. В общем случае он состоит из четырех частей, разделитель – точка с запятой:

- часть 1 – положительные числа,
- часть 2 – отрицательные числа,
- часть 3 – нули,
- часть 4 – текст.

В частном случае, если в шаблоне используется одна часть, то это отображение любых чисел, если две части, то первая для неотрицательных чисел, а вторая для отрицательных, если три части, то формат для текста общий.

Некоторые символы, используемые в шаблоне:

- 0 – одна цифра, в том числе незначащий нуль,
- # – одна цифра,
- ? – одна цифра с выравниванием по разделителю,
- _ – пробел (или как любой текст " "),
- @ – любое число текстовых символов.

Примеры пользовательских форматов в сравнении с общим форматом представлены на рис. 2.

Общий формат	Пользовательский формат			
	000,000	###,###	???,???	### " ?/5
42,1	042,100	42,1	42,1	42 1/5
23,76	023,760	23,76	23,76	23 4/5

Рис. 2. Пользовательский формат

Лабораторная работа №1. ПОСТРОЕНИЕ ГРАФИКОВ

1. Создание графиков

1.1. Запустить программу «Statistica v.6.0».

1.2. Создать новый файл командой «New» из меню «File».

1.3. Установить «Number of variables» (число переменных) 3, а «Number of cases» (число столбцов) 3 (рис. 3). Остальные параметры установить так, как указано на рис. 3.

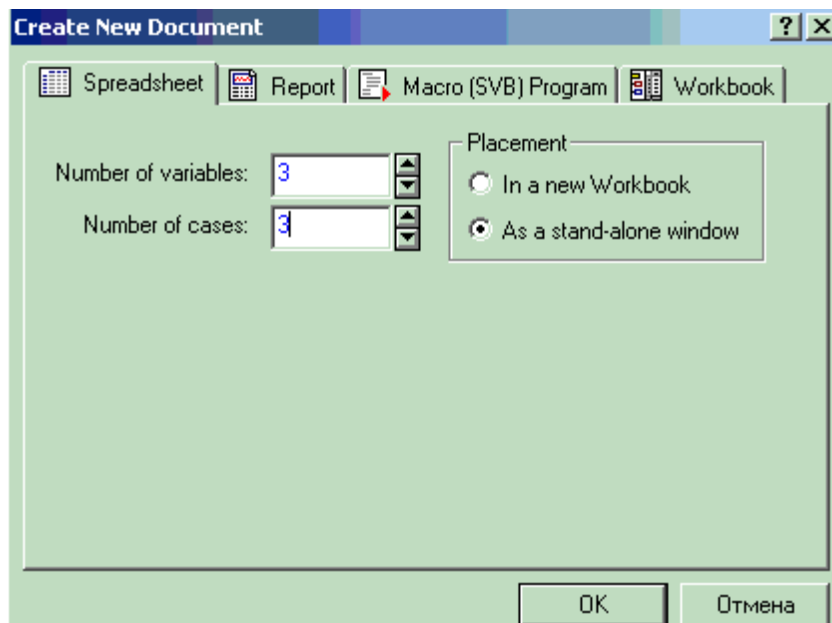


Рис. 3. Создание нового файла

1.4. Ввести, форматировать и сохранить данные:

1.4.1. Выполнить двойной щелчок по ячейке «Var1» и в открывшемся окне ввести «Name» 2002 (рис. 4).

1.4.2. Аналогично в ячейку «Var2» ввести значение 2003, а в ячейку «Var3» ввести 2004 (рис. 5).

1.4.3. Командой «Save as» из меню «File» присвоить имя файлу и сохранить в своей рабочей папке.

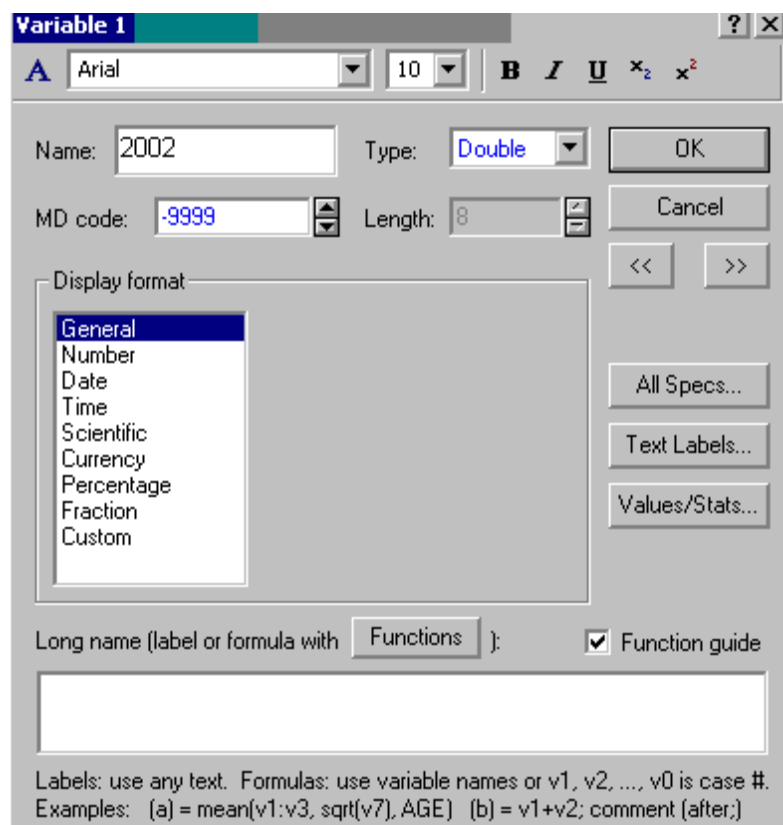


Рис. 4. Ввод имен переменных

	1	2	3
	2002	2003	2004
1			
2			
3			

Рис. 5. Результат ввода имен переменных

1.4.4. Выполнить двойной щелчок по ячейке «1» и ввести значение «дальнее зарубежье» (рис. 6). Нажать «Enter» и в ячейке «2» ввести «СНГ», вновь нажать «Enter» и в ячейке «3» ввести значение «Балтия».

1.4.5. В остальные ячейки ввести значения, приведенные на рис. 7.

	1 2002	2 2003	3 2004
дальнее зарубежье			
СНГ			
Балтия			

Рис. 6. Результат ввода имен cases

Регион	Экспорт из Свердловской области		
	1 2002	2 2003	3 2004
дальнее зарубежье	2621,39	3014,6	3316,06
СНГ	309,06	355,42	390,96
Балтия	30,96	34,06	37,46

Рис. 7. Ввод значений переменных

1.4.6. Щелчком правой клавиши мыши вызвать контекстное меню в левой верхней ячейке. Выбрать «Format Info Box Text».

1.4.7. Выбрать шрифт (рис. 8) и выравнивание (рис. 9).

1.4.8. Аналогично командой «Format Header» форматировать заголовок в верхней строке по центру, шрифт «Arial» 10 пт жирный (см. рис. 7). Сохранить данные в файле с расширением .STA.

1.5. Построить график.

1.5.1. Выполнить команду «2D Graphs /Line Plots (Variables)» из меню «Graphs».

1.5.2. Нажать кнопку «Variables» на вкладке «Quick». Выбрать все переменные (рис.10). Выбрать тип графика «Multiple».

1.5.3. На вкладке «Options 1» в списке «Case labels» выбрать значение «Case names». В строку «Custom title» ввести заголовок «Экспорт

из Свердловской области», нажать «OK». Результат представлен в прил. 1.
Сохранить график в файле с расширением .STW.

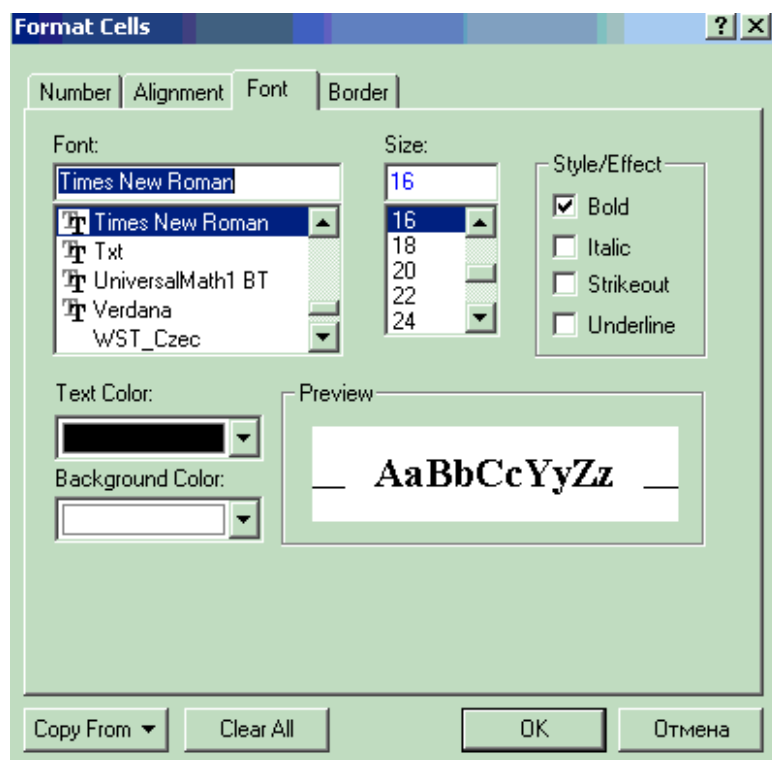


Рис. 8. Выбор шрифта

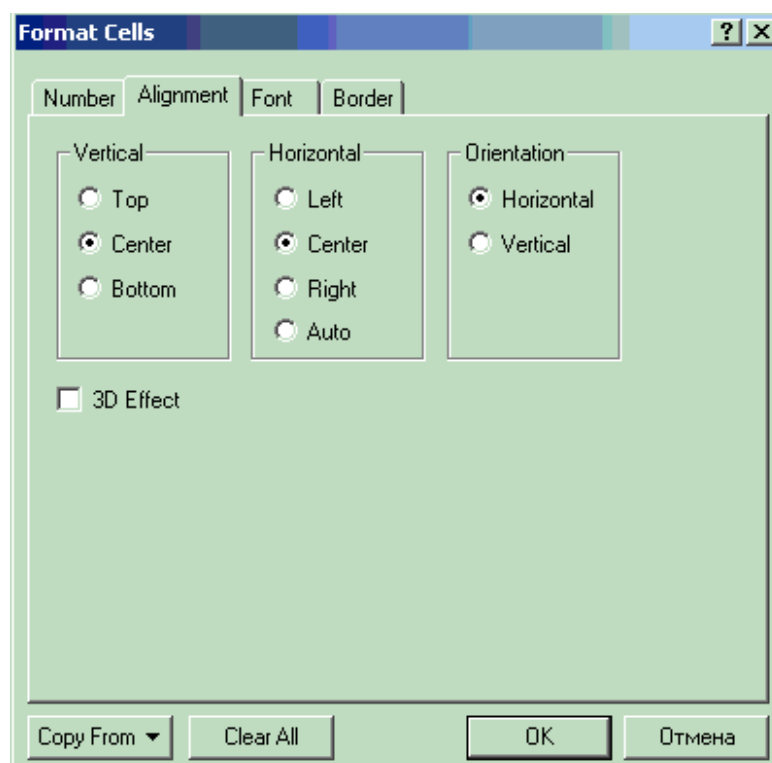


Рис. 9. Установка выравнивания

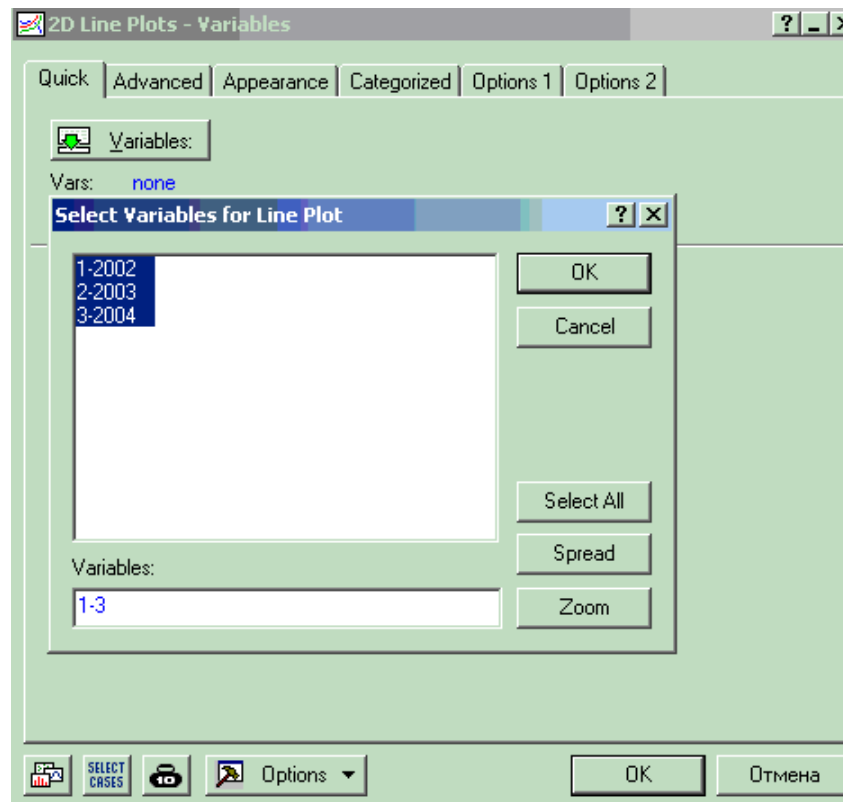


Рис. 10. Выбор переменных

1.6. Генерировать другой вид графика. Для этого произвести следующие действия:

1.6.1. Нажать кнопку «2D Line Plots – Variables» в левом нижнем углу окна программы (при ее отсутствии повторить п.1.5.1).

1.6.2. В левом нижнем углу открывшегося окна нажать первую кнопку «Graphs Gallery».

1.6.3. В окне «Graphs Gallery» выбрать вид графика «2D Graphs / / Line Plots (Cases Profiles)», нажать «ОК», вновь выбрать все переменные, «ОК». Результат представлен в прил. 2. Сохранить график под новым именем.

1.7. Добавить для кривой линии «Дальнее зарубежье» диаграмму, состоящую из столбцов:

1.7.1. Щелкнуть правой клавишей мыши по кривой «Дальнее зарубежье».

1.7.2. В контекстном меню выбрать «Bar Options».

1.7.3. В окне «Bars» включить переключатель «Display bars» и установить толщину столбца диаграммы «Width» (рис. 11), «OK». Результат представлен в прил. 3. Сохранить график под новым именем.

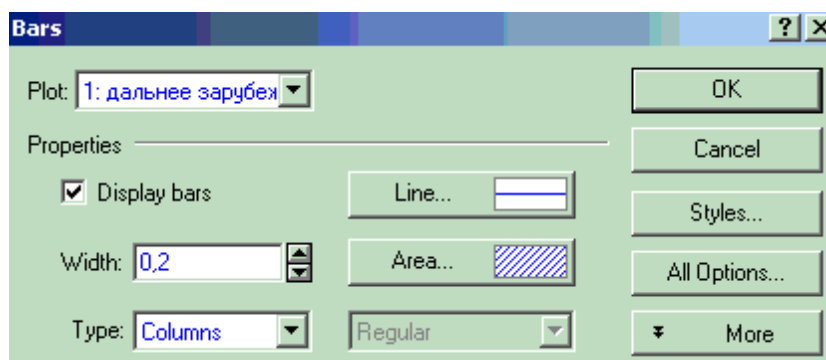


Рис. 11. Толщина столбцов

1.8. Закрыть все графики.

1.9. Создать график с двумя осями ординат:

1.9.1. Для изменения вида исходных данных добавить пустые строки, выделив последнюю строку и выбрав в контекстном меню (правая клавиша мыши) команду «Add Cases»; при появлении запроса о закрытии предыдущего анализа выбрать вариант «Да».

1.9.2. Ввести число добавляемых строк и место их расположения (рис. 12).

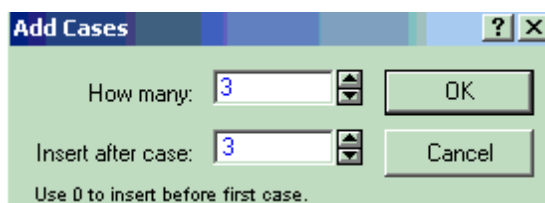


Рис. 12. Добавление строк

1.9.3. Вырезать поочередно заголовки строк (Cases): «дальнее зарубежье», «СНГ», «Балтия» командой «Cut» из контекстного меню и вставить их в 3 ячейки последней пустой строки командой «Paste» для временного хранения.

1.9.4. Передвинуть мышью числовые данные, используя для временного хранения ячейки остальных двух пустых строк. Сдвиг аналогичен перетаскиванию в Excel (рис. 13).


1.9.5. Выделить текст «Дальнее зарубежье» и вырезать в буфер.

1.9.6. Двойным щелчком по заголовку первого столбца данных вызвать окно настройки переменной (см. рис. 4)

1.9.7. В строке «Name» вместо 2002 вставить из буфера «Дальнее зарубежье», «ОК».

1.9.8. Аналогично изменить имена второй и третьей переменных (см. рис. 13).

1.9.9. Изменить заголовок столбца «Cases» на «Год». Ввести заголовки строк (см. рис. 13). Удалить пустые строки с помощью контекстного меню командой «Delete cases». Сохранить файл под новым именем в рабочей папке.



Год	Экспорт из Свердловской области		
	1	2	3
	дальнее зарубежье	СНГ	Балтия
2002	2621,39	309,06	30,96
2003	3014,6	355,42	34,06
2004	3316,06	390,96	37,46

Рис. 13. Результат транспонирования строк и столбцов

1.9.10. Выполнить команду «2D Graphs / Line Plots (Variables)» из меню «Graphs».

1.9.11. На вкладке «Advanced» выбрать тип графика «Double Y».

1.9.12. Нажать кнопку «Variables» на этой же вкладке. Выбрать переменные для левой и правой осей Y (рис. 14).

1.9.13. Закрыть окно, нажав ОК. Результат представлен в прил. 4. Сохранить график под новым именем.

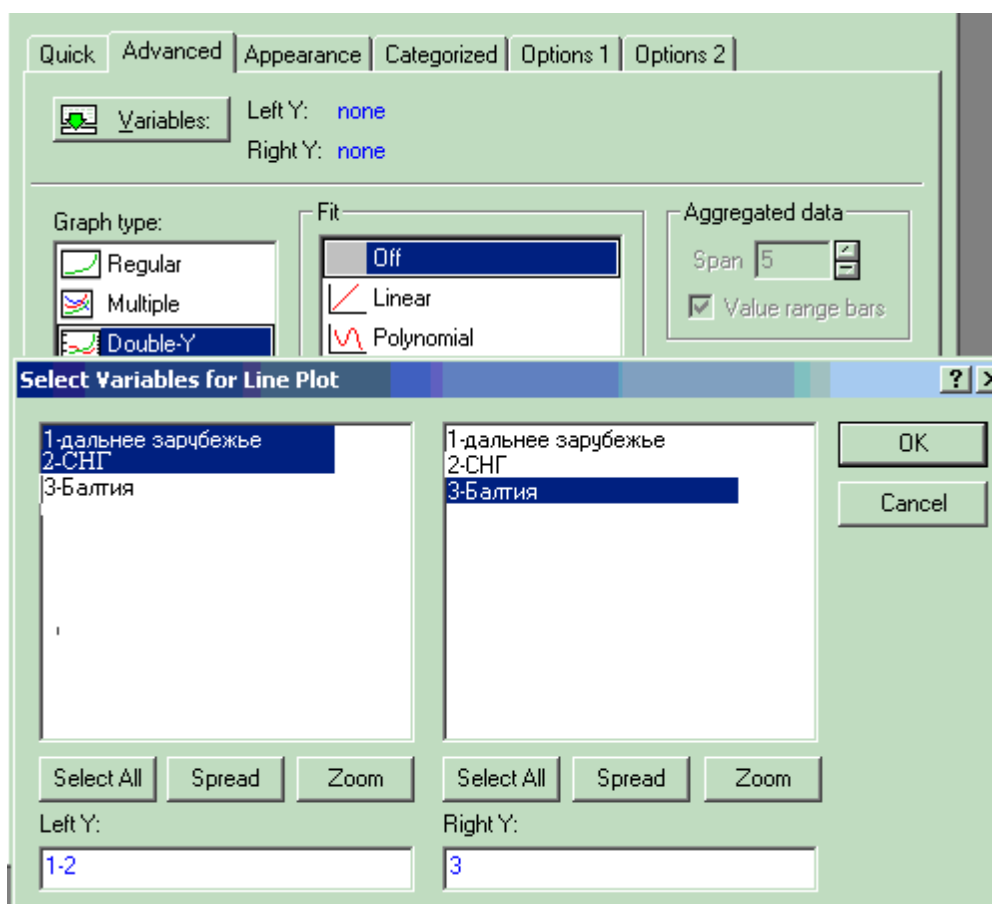


Рис. 14. Выбор типа графика

2. Форматирование графиков

2.1. Изменить тип и толщину кривой «Балтия», щелкнув по ней правой клавишей мыши и выбрав команду «Properties»:

2.1.1. В окне «General» щелкнуть по кнопке «Line».

2.1.2. Выбрать непрерывную линию толщиной 2 болотного цвета (рис. 15). Нажать «Close».

2.1.3. В окне «General» щелкнуть по кнопке «Markers».

2.1.4. Выбрать форму маркера круг, размер 8. Нажать «Close», «OK».

2.2. Аналогично изменить кривые «СНГ», «Дальнее зарубежье». Цвет линий не менять (прил. 5).

2.3. Контекстной командой «Graph Properties (All Options)», вызванной на фоне подложки, изменить параметры графика:

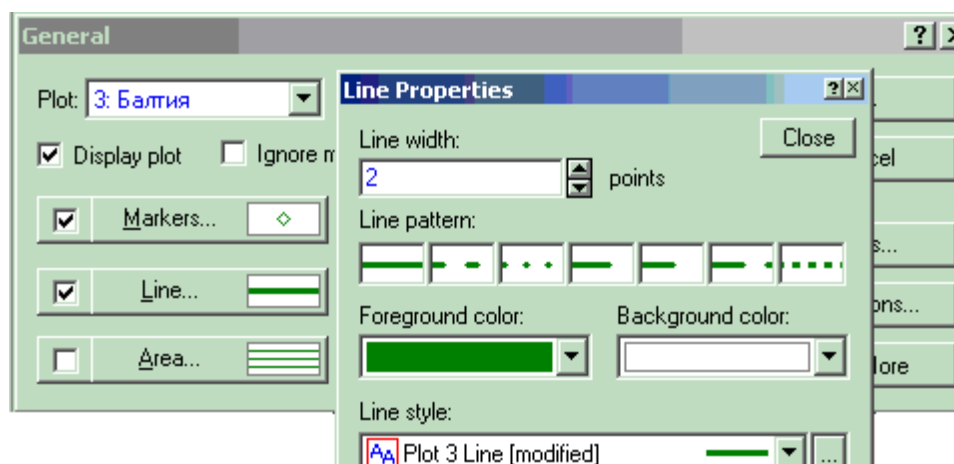


Рис. 15. Окно форматирования

2.3.1. На вкладке «Graph Window» нажатием кнопки «Outside Background Color» изменить цвет подложки на голубой.

2.3.2. На вкладке «Graph Titles / Text» изменить заголовок на «Экспорт из Свердловской области» (см. прил. 5). Шрифт «Arial» 14 пт жирный.

2.3.3. На вкладке «Plot: Point Labels» установить параметры, указанные на рис. 16. Нажатием кнопки «Font» изменить шрифт на «Times New Roman» 11 пт. При этом установку произвести для трех переменных, т.е. изменять поочередно значения в раскрывающемся списке «Plot».

2.3.4. На вкладке «Axis Titles» ввести название оси X, цвет шрифта вишневый, остальные параметры указаны на рис. 17. Аналогично ввести названия левой и правой оси Y (см. прил. 5).

2.3.5. На вкладке «Axis Scaling» изменить диапазон левой оси Y (рис. 18), затем нажатием кнопки «Edit step» изменить цену деления этой оси (рис. 19). Аналогично изменить правую ось Y (см. прил. 5).

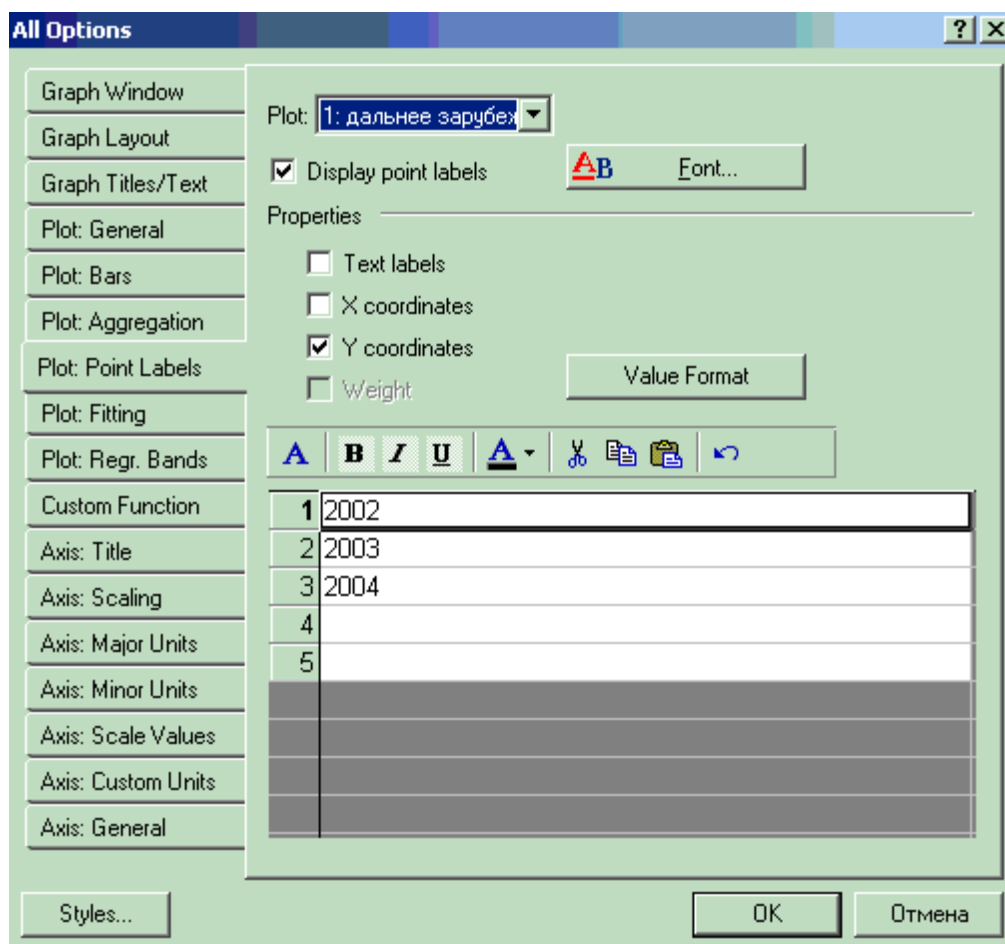


Рис. 16. Вкладка «Plot: Point Labels»

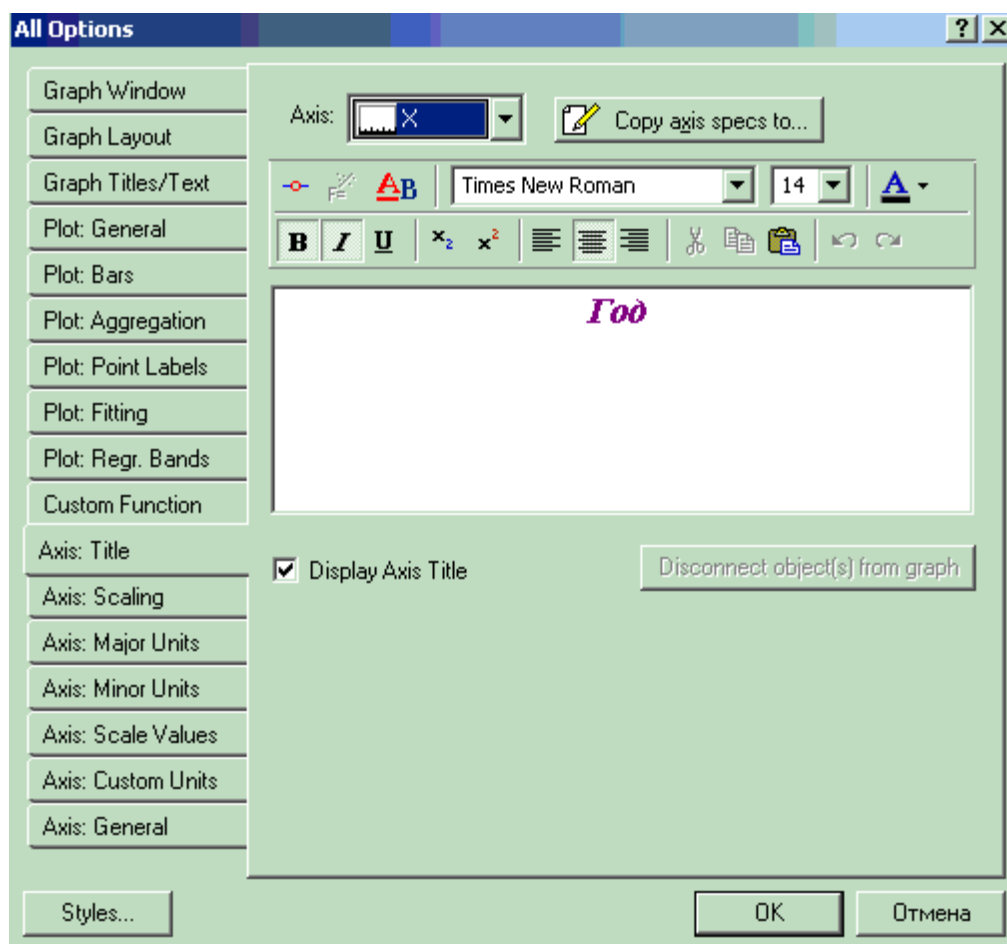


Рис. 17. Вкладка «Axis: Title»

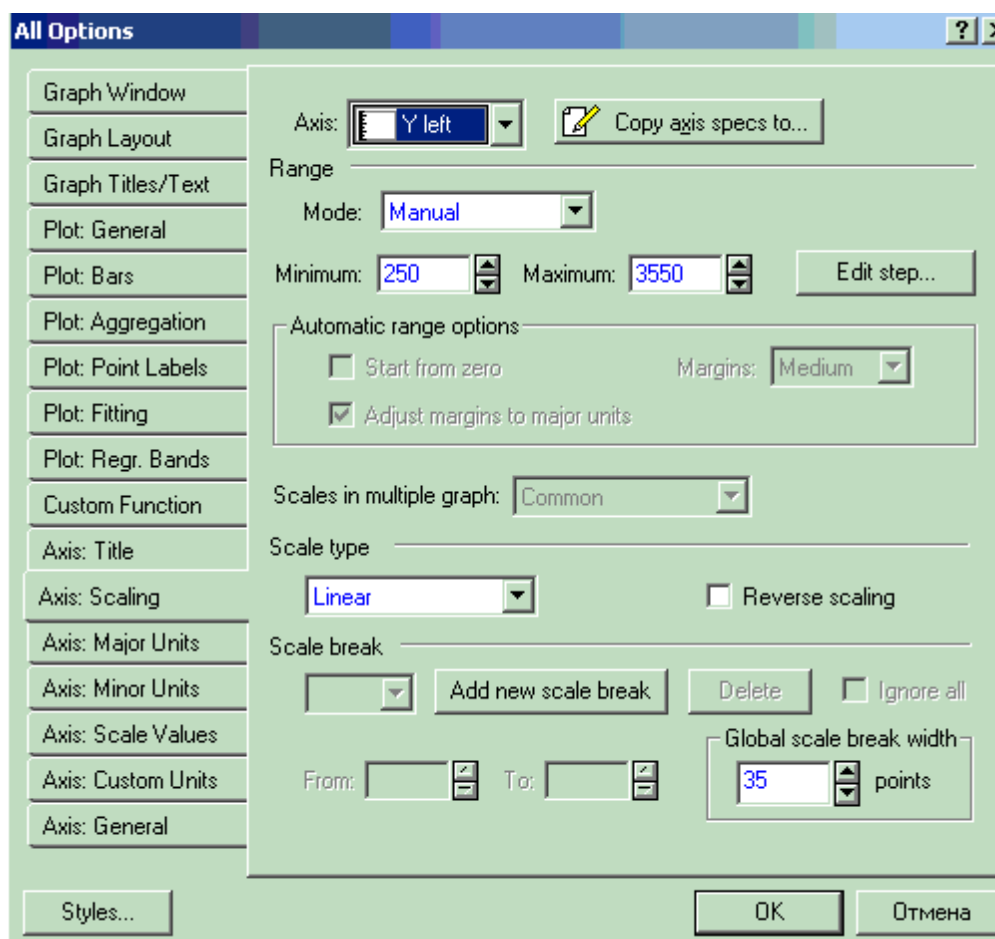


Рис. 18. Вкладка «Axis: Scaling»

2.3.6. На вкладке «Axis Major Units» нажатием кнопки «Gridlines» изменить тип линии сетки левой оси Y и оси X. Выбрать второй слева образец (равномерная штриховая) (см. прил. 5).

2.3.7. На вкладке «Axis Scale Values» включить опцию «Automatic – at major tickmarks» и выключить опцию «Data Values:» для левой и правой оси Y, «OK» (см. прил. 5).

2.3.8. Если расположение кривой переменной «СНГ» отличается от представленного в прил. 5, то выполнить контекстную команду «Properties», выполненную на фоне этой кривой, далее щелчок по кнопке «All Options...», установить значение «Right» в опции «Assignment of axis».

2.3.9. Контекстной командой «Rename», вызванной на фоне значка текущего графика в древовидной иерархической структуре, расположенной

слева от графика, изменить имя рабочего листа на «внешнеэкономическая деятельность» (см. прил. 5).

2.4. Сохранить график под новым именем. Закрыть все файлы.

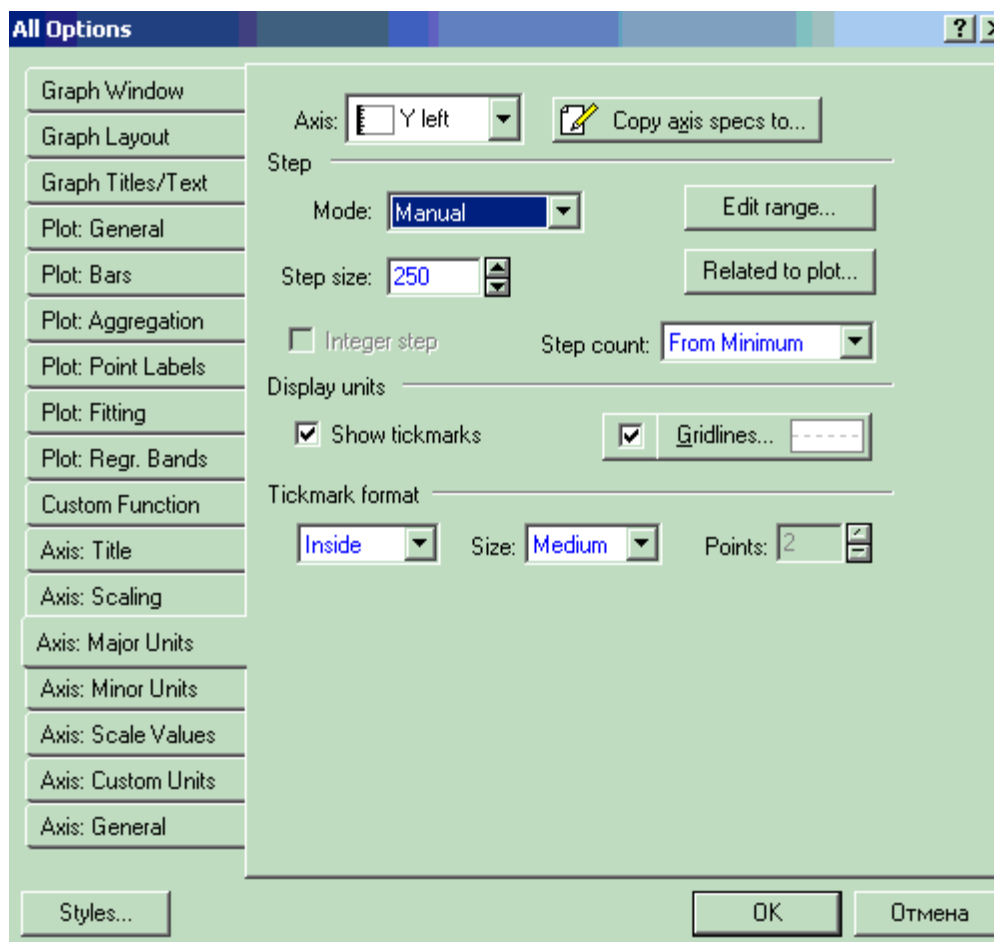


Рис. 19. Цена деления левой оси Y

3. График биржевых цен

3.1. Для сравнения построить в Excel и в Statistica биржевой график цен фьючерсных контрактов на казначейские облигации в тридцать вторых долях доллара, имея данные за 5 дней.

3.1.1. Построить график в Excel:

3.1.1.1. Ввести и форматировать данные в Excel (рис. 20). Для ввода тридцать вторых долей создать новый пользовательский дробный формат.

	A	B	C	D	E
1	Дата	Open	High	Low	Close
2	10.12.2004	117 15/32	117 22/32	117 12/32	117 14/32
3	11.12.2004	118 19/32	118 21/32	118 19/32	118 19/32
4	12.12.2004	119 4/32	119 15/32	119 4/32	119 10/32
5	15.12.2004	119 7/32	119 7/32	119 3/32	119 3/32
6	16.12.2004	119 3/32	119 9/32	119	119 5/32
7					
8					

Рис. 20. Исходные данные в дробном формате

3.1.1.2. Построить график. Тип «Биржевая», вид «Биржевая диаграмма для наборов из четырех значений (курс открытия, курс закрытия, самый высокий курс, самый низкий курс)». Включить опцию «Категории» по оси X на вкладке «Оси» для автоматического исключения выходных дней. Ввести заголовок графика и заголовки осей. Убрать легенду и линии сетки.

3.1.1.3. Форматировать график. Изменить размер графика. Изменить и толщину осей, а также шрифт значений и названий, шрифт заголовка. Выбрать цвет: фон графика голубой, заголовок коричневый, названия осей вишневые (прил. 6).

3.1.1.4. Подготовить данные для загрузки в Statistica: копировать данные с транспонированием с листа 1 на лист 2 (рис. 21).

3.1.1.5. Сохранить файл. Закрыть Excel.

3.1.2. Построить график в Statistica:

3.1.2.1. Командой «Open» из меню «File» открыть в Statistica созданный в Excel файл.

3.1.2.2. В окне «Select Sheet to Import» выбрать «Лист2».

	A	B	C	D	E	F	
1	Дата	10.12.2004	11.12.2004	12.12.2004	15.12.2004	16.12.2004	
2	Open	117 15/32	118 19/32	119 4/32	119 7/32	119 3/32	
3	High	117 22/32	118 21/32	119 15/32	119 7/32	119 9/32	
4	Low	117 12/32	118 19/32	119 4/32	119 3/32	119	
5	Close	117 14/32	118 19/32	119 10/32	119 3/32	119 5/32	
6							

Рис. 21. Транспонированные данные

3.1.2.3. Установить опции выбора переменных и Cases (рис. 22).

Результат импорта представлен на рис. 23.

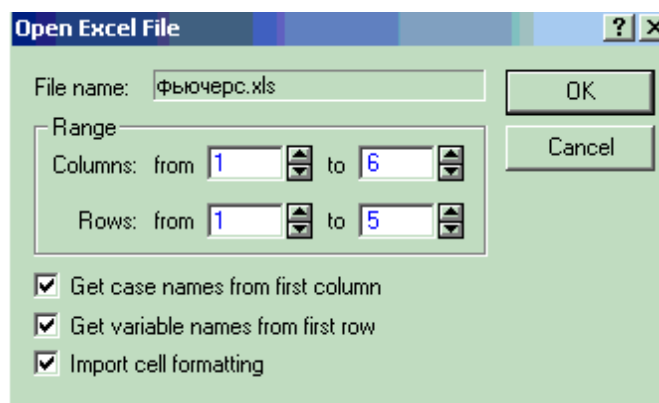


Рис. 22. Выбор переменных

Data: фьючерс* (5v by 4c)					
	1	2	3	4	5
	12/10/04	12/11/04	12/12/04	12/15/04	12/16/04
Open	117 15/32	118 19/32	119 4/32	119 7/32	119 3/32
High	117 22/32	118 21/32	119 15/32	119 7/32	119 9/32
Low	117 12/32	118 19/32	119 4/32	119 3/32	119
Close	117 14/32	118 19/32	119 10/32	119 3/32	119 5/32

Рис. 23. Импорт данных

3.1.2.4. Изменить имена переменных, поменять местами день и месяц (см. п.1.4.1). Результат представлен на рис. 24.

Data: фьючерс* (5v by 4c)					
	1	2	3	4	2 5
	10/12/04	11/12/04	12/12/04	15/12/04	16/12/04
Open	117 15/32	118 19/32	119 4/32	119 7/32	119 3/32
High	117 22/32	118 21/32	119 15/32	119 7/32	119 9/32
Low	117 12/32	118 19/32	119 4/32	119 3/32	119
Close	117 14/32	118 19/32	119 10/32	119 3/32	119 5/32

Рис. 24. Исходные данные

3.1.2.5. Выполнить команду «2D Graphs / Box Plots» из меню «Graphs».

3.1.2.6. Выбрать тип графика «Box Whiskers», составной «Multiple».

3.1.2.7. Нажать кнопку «Variables» на вкладке «Quick». Выбрать все переменные в левой части, справа не выбирать переменные, «OK», «OK».

3.1.2.8. Контекстной командой «Axis Properties» изменить формат оси Y: на вкладке «Scale Values» ввести шаблон пользовательского формата, аналогичный формату, созданному в Excel (рис. 25). Результат представлен в прил. 7. Сохранить и закрыть файл.

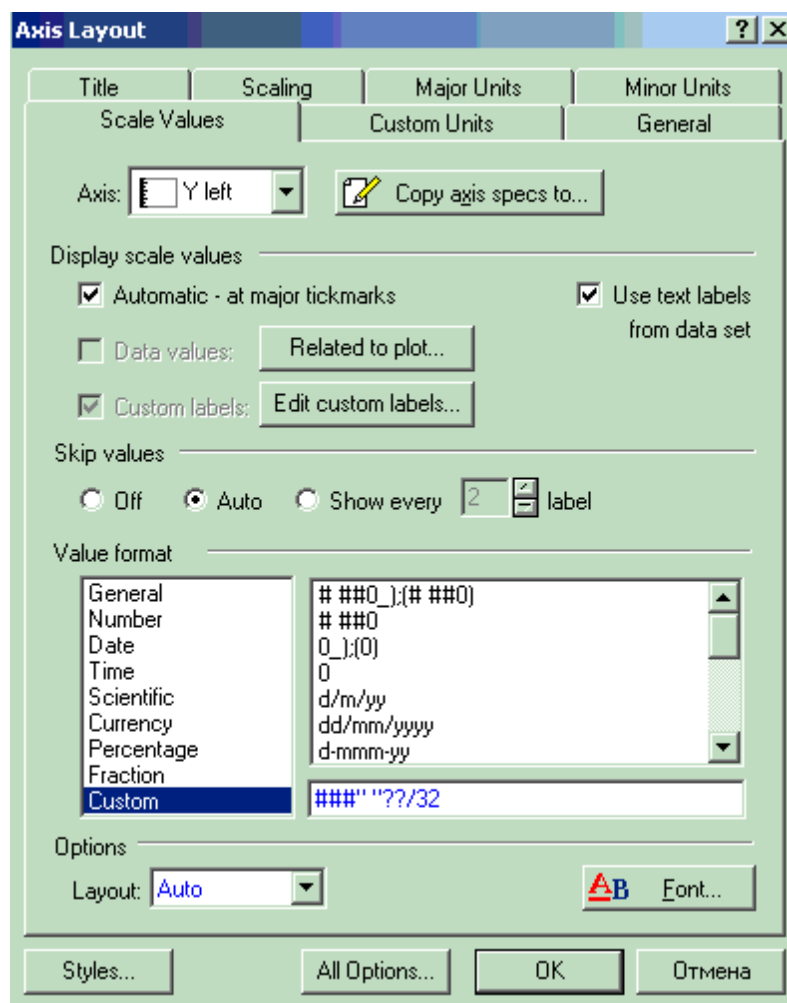


Рис. 25. Пользовательский формат

Лабораторная работа №2. РЕГРЕССИОННЫЙ АНАЛИЗ

1. Множественная линейная регрессия

1.1. Создать новый файл, в котором будут использованы исходные данные, приведенные в прил. 8:

1.1.1. Командой «All Variable Specs» из меню «Data» ввести переменные: объем продаж за текущий месяц S , объем продаж за предыдущий месяц SI , расходы на рекламу в текущем месяце A , расходы на рекламу в предыдущем месяце AI , число туристов в текущем месяце T , средняя температура воздуха W , индекс розничных цен в текущем месяце P (см. прил. 8). Имена переменных, например S , вводить в столбец «Name», а комментарий, например «объем продаж за текущий месяц», в столбец «Long Name (label or formula)»

1.1.2. Ввести число наблюдений за 24 месяца, значения переменных, заголовки (см. прил. 8).

1.2. Запустить множественный регрессионный анализ:

1.2.1. Выполнить команду «Multiple Regression» из меню «Statistics», вкладка «Advanced», кнопка «Variables».

1.2.2. Выбрать зависимые (dependent) и независимые (independent) переменные (рис. 26), «ОК», «ОК».

1.2.3. Результаты стандартного множественного регрессионного анализа представлены в прил. 9.

1.2.4. Нажать кнопку «Cancel» и изменить процедуру на пошаговую «Stepwise» (рис. 27), «ОК».

1.2.5. Выбрать метод на вкладке «Advanced» (рис. 28).

1.2.6. Выбрать параметры пошаговой процедуры (рис. 29), «ОК». Результат представлен в прил. 10.

1.2.7. Щелкнуть по кнопке «Summary: Regression Results» на вкладке «Advanced» (см. прил. 10).

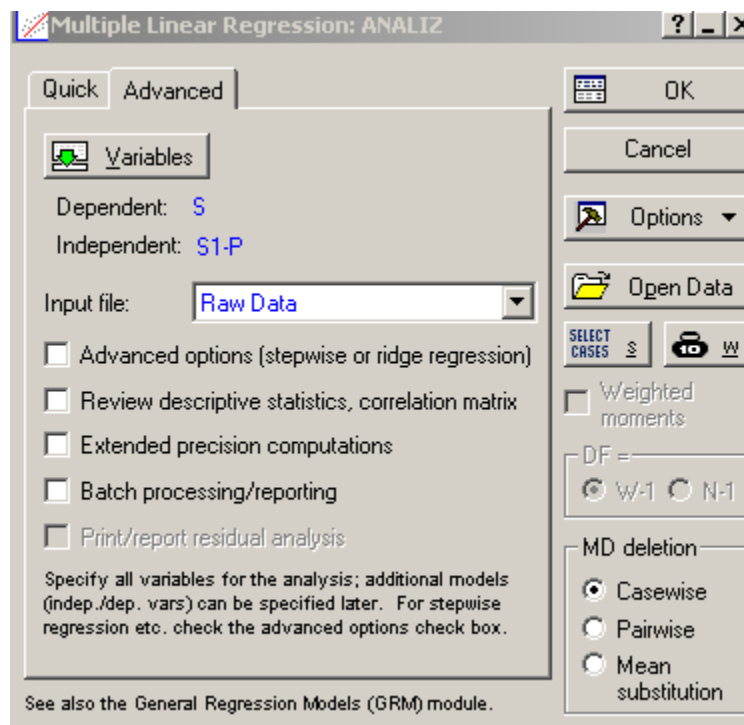


Рис. 26. Окно множественной регрессии

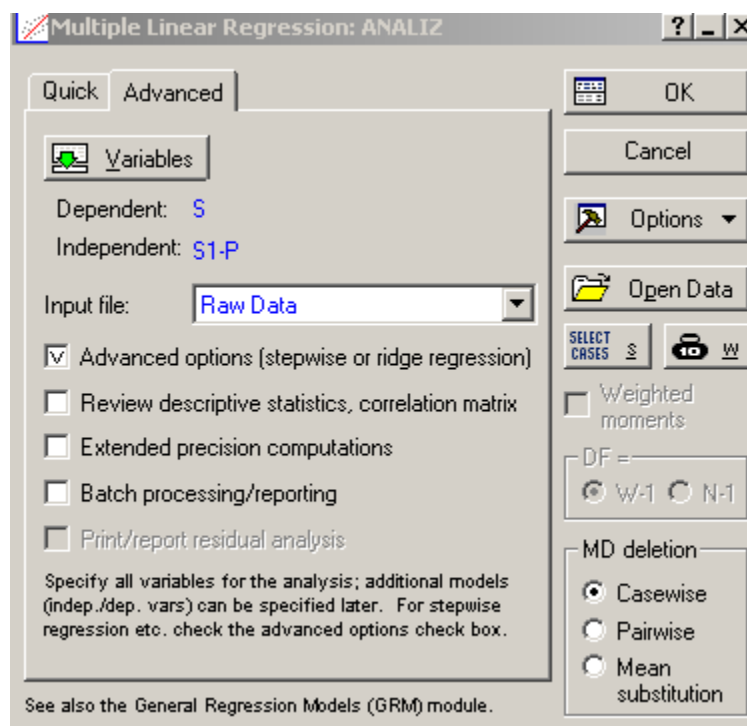


Рис. 27. Пошаговый анализ

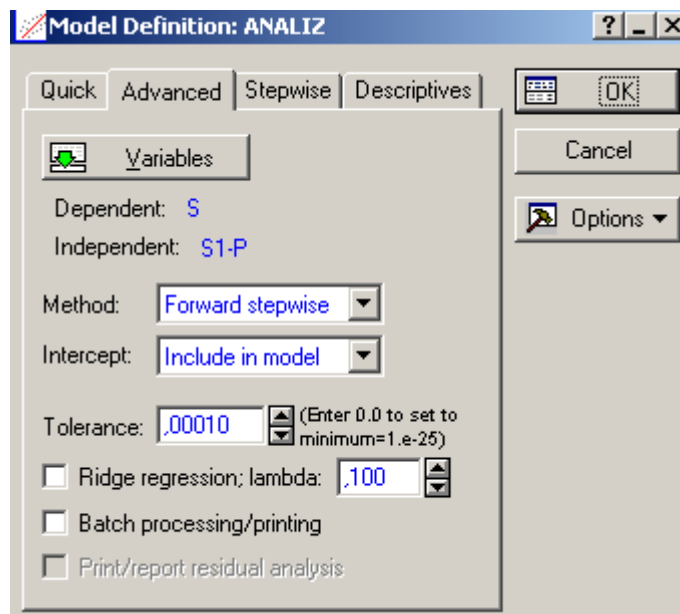


Рис. 28. Выбор метода

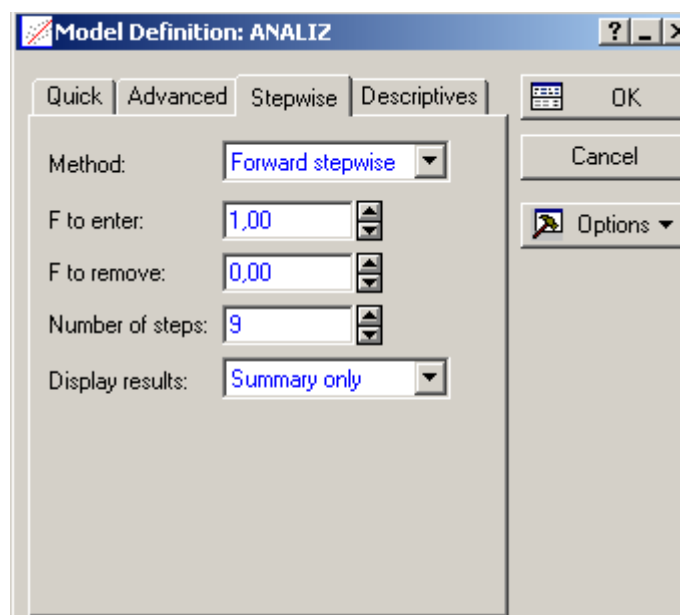


Рис. 29. Параметры пошаговой процедуры

1.2.8. Итоги представлены в прил. 11.

1.2.9. Вернуться в окно анализа, щелкнув по кнопке «Multiple Regression...».

1.2.10. Произвести анализ остатков, щелкнув по кнопке «Perform residual analysis» на вкладке «Residuals/assumptions/prediction». Результат представлен в прил. 12.

1.2.11. Исследовать зависимость между остатками и прогнозом, щелкнув по кнопке «Durbin-Watson statistic» на вкладке «Advanced». Результат представлен в прил. 13.

1.2.12. Вернуться в окно анализа, щелкнув по кнопке «Residual Analysis:...».

1.2.13. Щелкнуть по кнопке «Predicted vs. Observed» на вкладке «Scatterplots». Результат представлен в прил. 14.

1.2.14. Вернуться в окно анализа, щелкнув по кнопке «Residual Analysis:...».

1.2.15. Щелкнуть по кнопке «Cancel» для того, чтобы вернуться в окно «Model definition» (см. рис. 28), далее «OK».

1.2.16. Щелкнуть по кнопке «Predict dependent variable» на вкладке «Residuals/assumptions/prediction».

1.2.17. Ввести данные для прогноза на следующий месяц (рис. 30), щелкнуть по кнопке «OK».

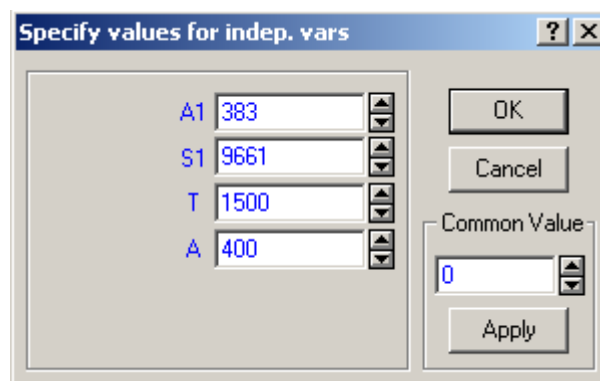
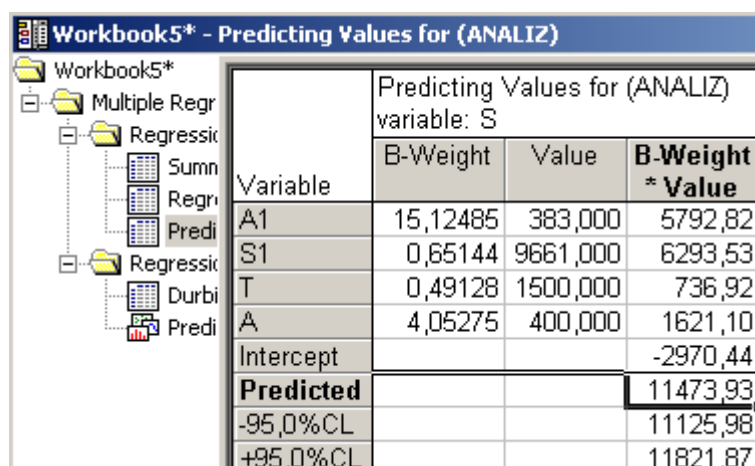


Рис. 30. Данные для прогноза

1.2.18. Результат представлен на рис. 31. Ожидаемый объем продаж 11473,93.



Predicting Values for (ANALIZ)			
variable: S			
Variable	B-Weight	Value	B-Weight * Value
A1	15,12485	383,000	5792,82
S1	0,65144	9661,000	6293,53
T	0,49128	1500,000	736,92
A	4,05275	400,000	1621,10
Intercept			-2970,44
Predicted			11473,93
-95,0%CL			11125,98
+95,0%CL			11821,87

Рис. 31. Ожидаемый объем продаж

1.3. Сохранить и закрыть все файлы.

2. Индивидуальное задание

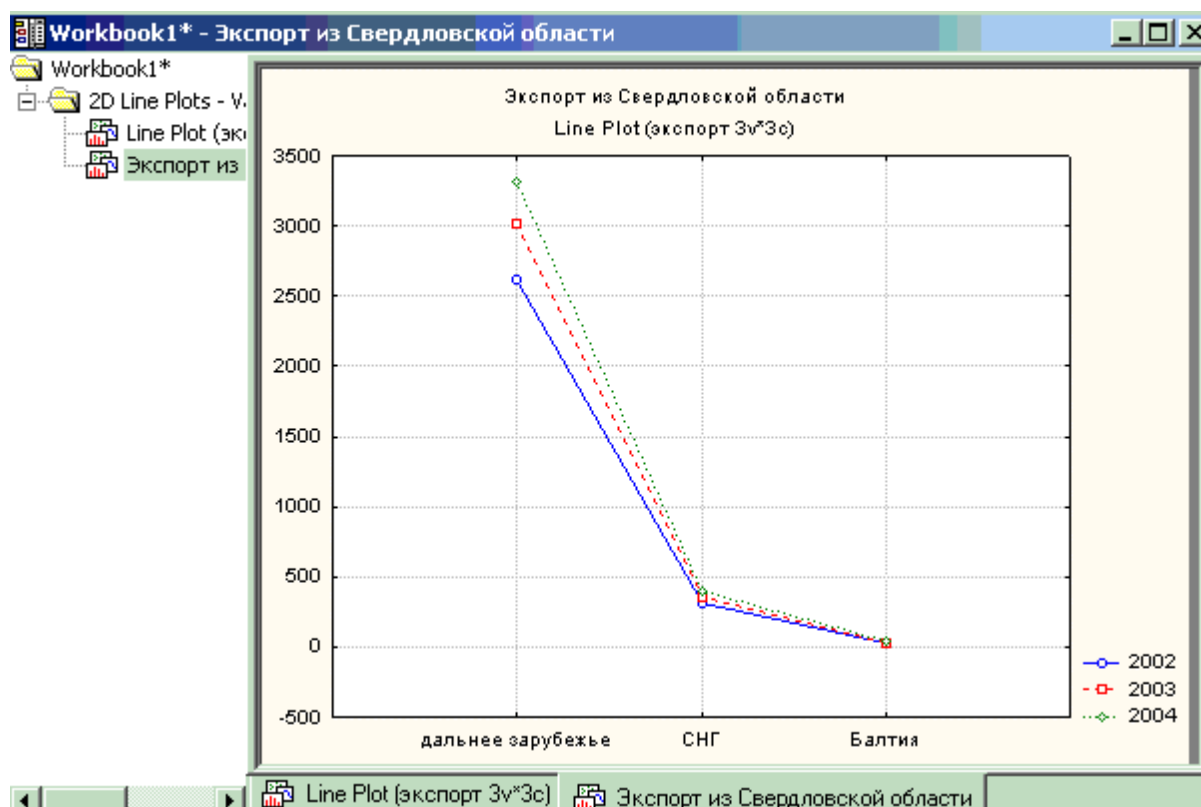
2.1. Создать новый файл. Исходные данные взять у преподавателя.

2.2. Выполнить анализ данных с помощью множественной регрессии.

СПИСОК ЛИТЕРАТУРЫ

1. Боровиков В.П. STATISTICA. Искусство анализа данных на компьютере: для профессионалов / В.П. Боровиков. – СПб.: Питер, 2003. – 688 с. : ил.
2. Макарова Н.В. Статистика в Excel: учебное пособие / Н.В. Макарова, В.Я. Трофимец. – М.: Финансы и статистика, 2002. – 368 с. : ил.
3. Додж Марк. Эффективная работа: Excel 2002 : Пер. с англ. / Марк Додж, Крейг Стинсон. – СПб.: Питер, 2003. – 992 с. : ил.
4. Руденко В.И. Статистика. Краткий курс: пособие студентам для подготовки к экзаменам / В.И. Руденко. – М.: Дашков и К°, 2005. – 188 с.

Предварительный график



Экспорт по годам

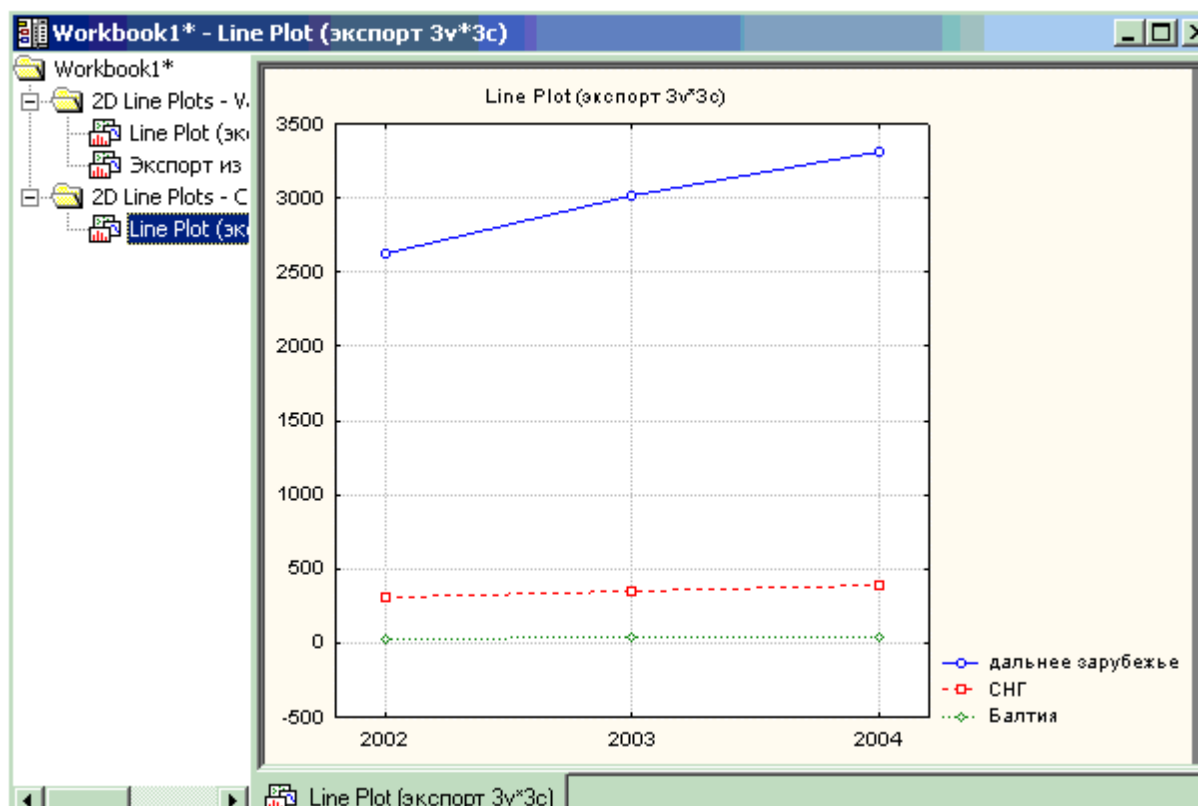


График со столбцами

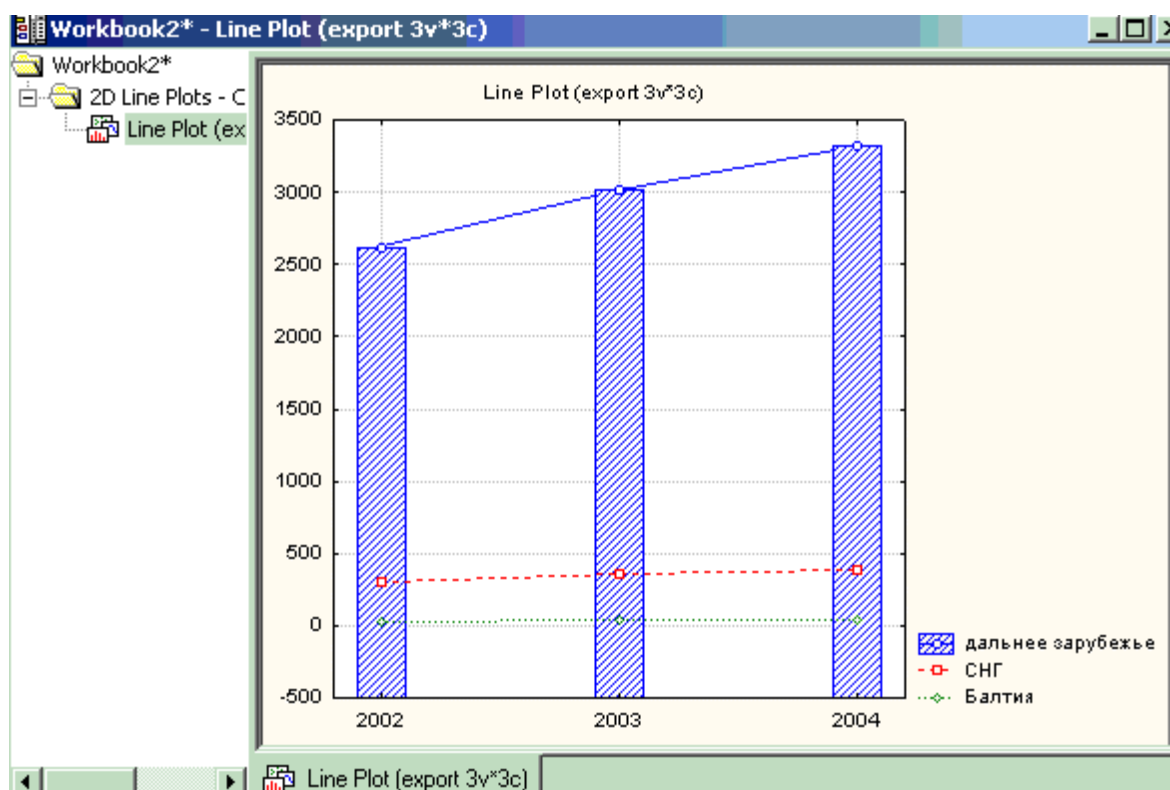
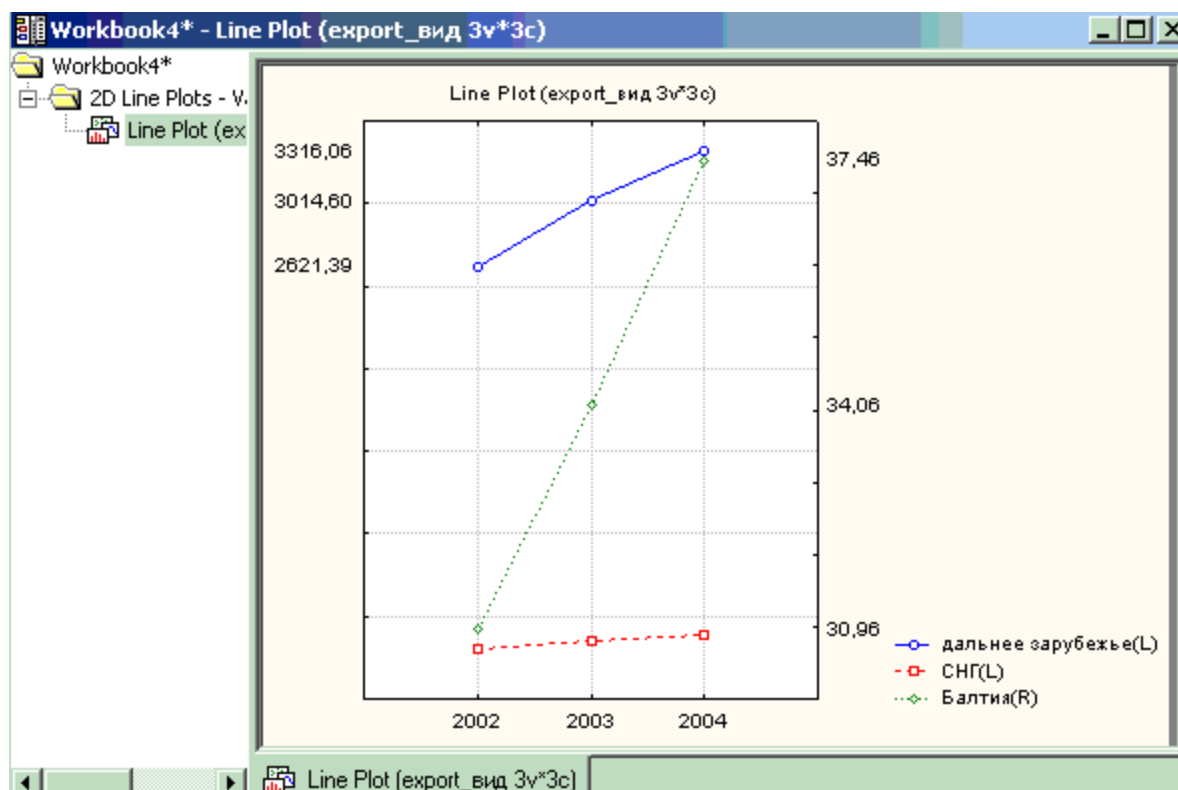
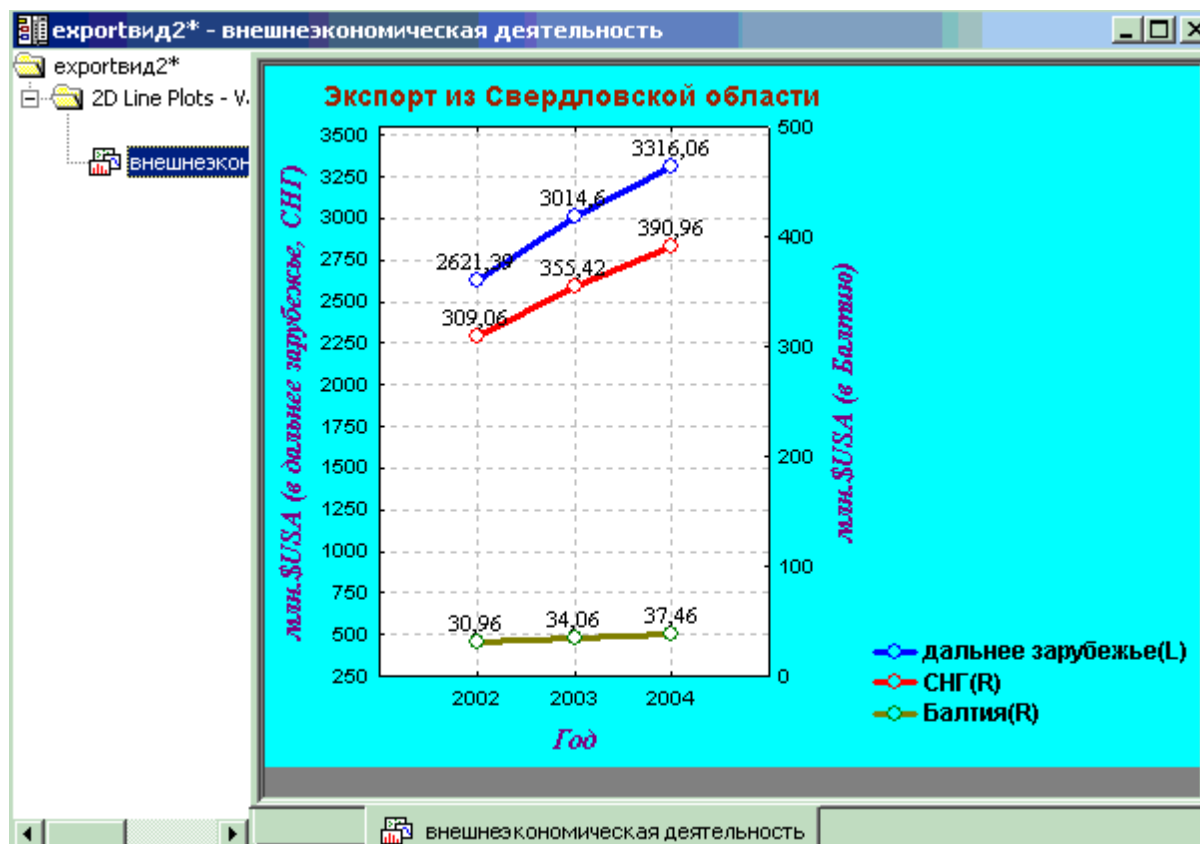


График с двумя шкалами по ординате

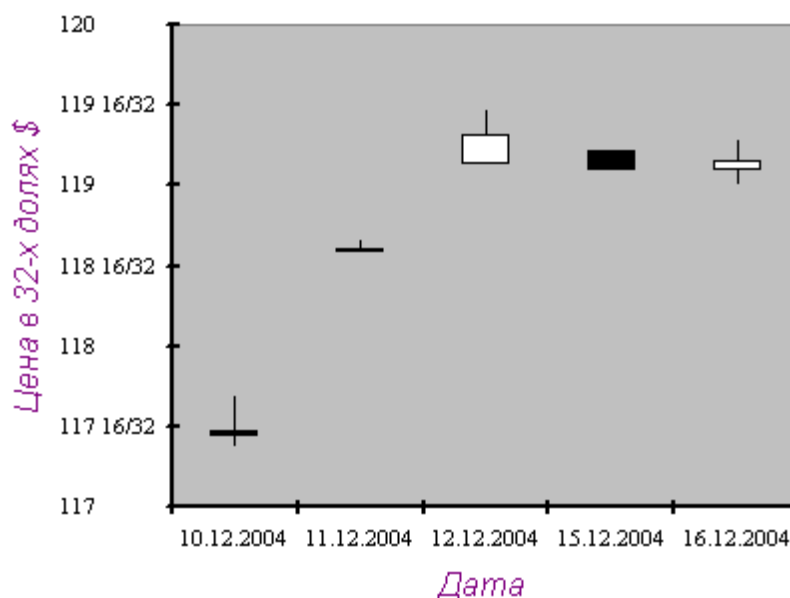


Приложение 5
График биржевых цен в Excel

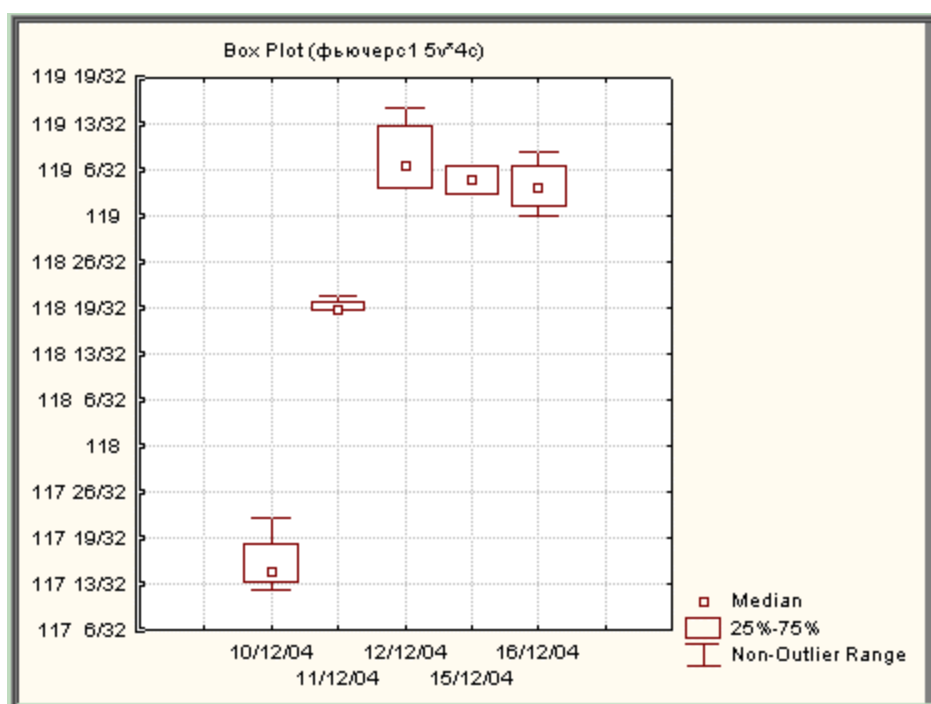


Приложение 6
Форматирование графика

Фьючерс на казначейские облигации



Приложение 7
График биржевых цен в Statistica

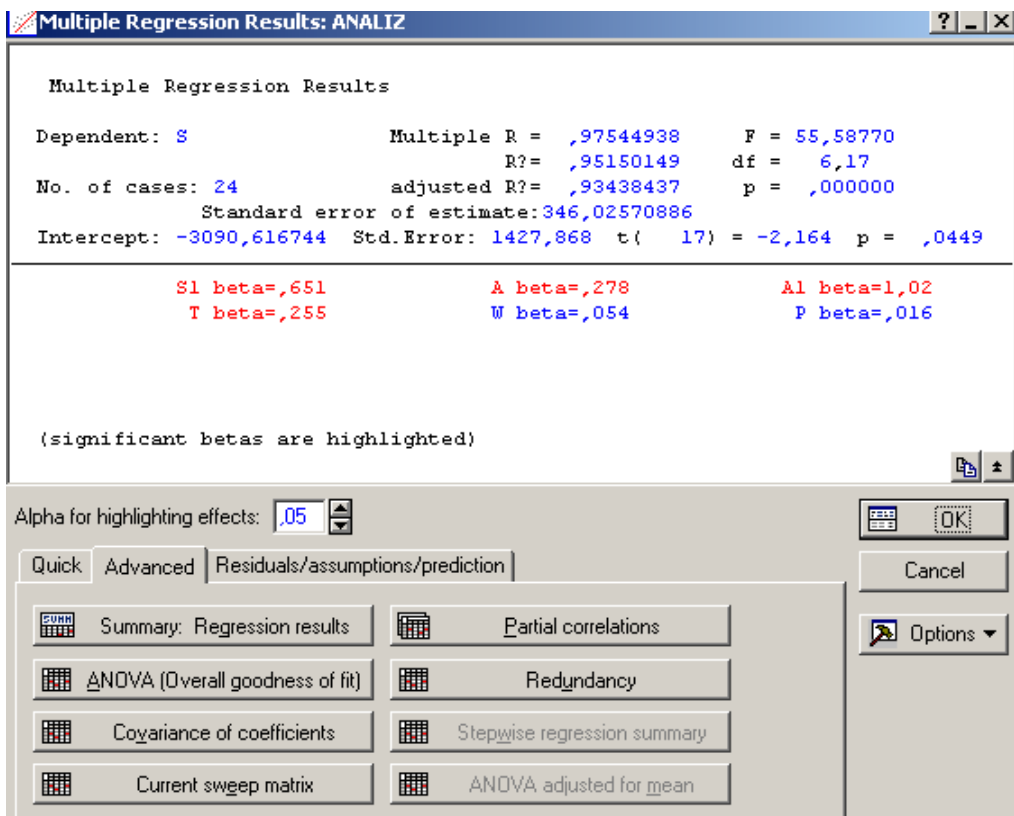


Исходные данные

Номер месяца	Объем продаж прохладительных напитков						
	1 S	2 S1	3 A	4 A1	5 T	6 W	7 P
1	8846,4	10132	288,5	200	1675	8,2	101,5
2	8774,7	8846,4	324,2	288,5	1405	9,4	103,8
3	9524,7	8774,7	332,2	324,2	723	15,1	109,2
4	11134,3	9524,7	439,2	332,2	2230	20,9	108,9
5	12239,8	11134,3	98,2	439,2	2615	25,7	112,4
6	8862,1	12239,8	212,8	98,2	3089	26,3	113,1
7	8646,9	8862,1	416	212,8	1741	22,8	114,1
8	11758,5	8646,9	327,7	416	2060	18,7	116
9	11867,2	11758,5	160,6	327,7	1777	15,5	116,6
10	9577,6	11867,2	403,1	160,6	1378,9	11,7	122,6
11	10898,4	9577,6	269,7	403,1	1253,3	5,6	119,5
12	9638,6	10898,4	280,5	269,7	794	6,5	130,6
13	9203,9	9638,6	335,1	280,5	1384,4	8,1	125
14	9231,1	9203,9	169,3	335,1	1392,5	10,3	124,2
15	7334,5	9231,1	206	169,3	2484,4	9,7	130,7
16	7647	7334,5	216,1	206	2777,5	22,4	131,6
17	7839,6	7647	322,2	216,1	3301,9	20,6	133,4
18	9787	7839,6	285,5	322,2	3635,9	26,8	139,1
19	9600	9787	79,2	285,5	3415,9	27,8	142,3
20	7199,9	9600	333,6	79,2	2606,8	18,3	139,9
21	9547,7	7199,9	293,1	333,6	2508	12	144,5
22	10187,5	9547,7	238,5	293,1	2834,1	9,7	143,9
23	9661,2	10187,5	255,4	238,5	2481,8	5,1	148
24	9189,2	9661,2	383,6	255,4	1474,4	2,8	149,3

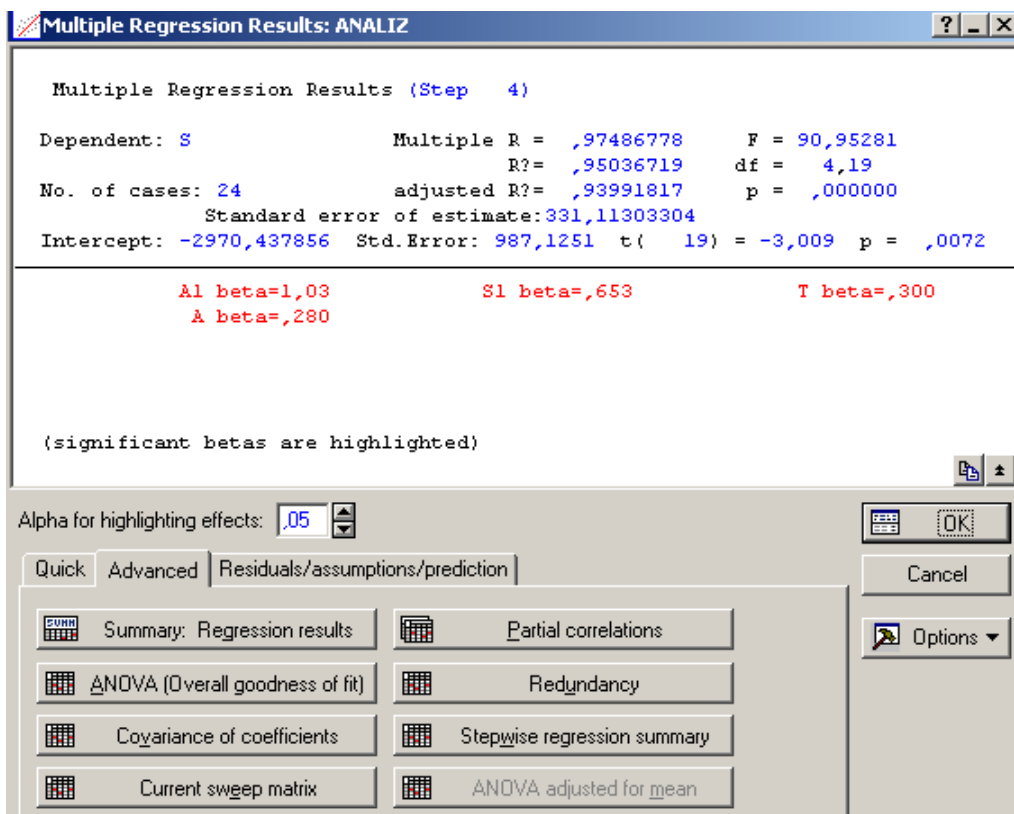
Приложение 9

Результат стандартной регрессии



Приложение 10

Результат пошаговой регрессии



Итоги пошаговой регрессии

Workbook5* - Regression Summary for Dependent Variable: S (ANALIZ)

Regression Summary for Dependent Variable: S (ANALIZ)
 R= ,97486778 R²= ,95036719 Adjusted R²= ,93991817
 F(4,19)=90,953 p<,00000 Std.Error of estimate: 331,11

	Beta	Std.Err. of Beta	B	Std.Err. of B	t(19)	p-level
N=24						
Intercept			-2970,44	987,1251	-3,00918	0,007214
A1	1,029462	0,058551	15,12	0,8602	17,58223	0,000000
S1	0,653376	0,059701	0,65	0,0595	10,94405	0,000000
T	0,299847	0,063616	0,49	0,1042	4,71341	0,000151
A	0,280419	0,063531	4,05	0,9182	4,41388	0,000298

Анализ остатков

Residual Analysis: ANALIZ

Dependent: S Multiple R : ,97486778 F = 90,95281
 R²: ,95036719 df = 4,19
 No. of cases: 24 adjusted R²: ,93991817 p = ,000000
 Standard error of estimate: 331,11303304
 Intercept: -2970,437856 Std. Error: 987,1251 t(19) = -3,009 p < ,0072

Quick Advanced Residuals Predicted Scatterplots Probability plots Outliers Save

☐ Histogram of residuals
☐ Casewise plot of residuals
☐ Residuals vs. independent var.
☐ Histogram of observed

Type of residual

☒ Raw residuals ☐ Deleted residuals
☐ Standard residuals ☐ Cook's distances
☐ Mahalanobis distances

Приложение 13

Итоги статистики Дарбина-Уотсона

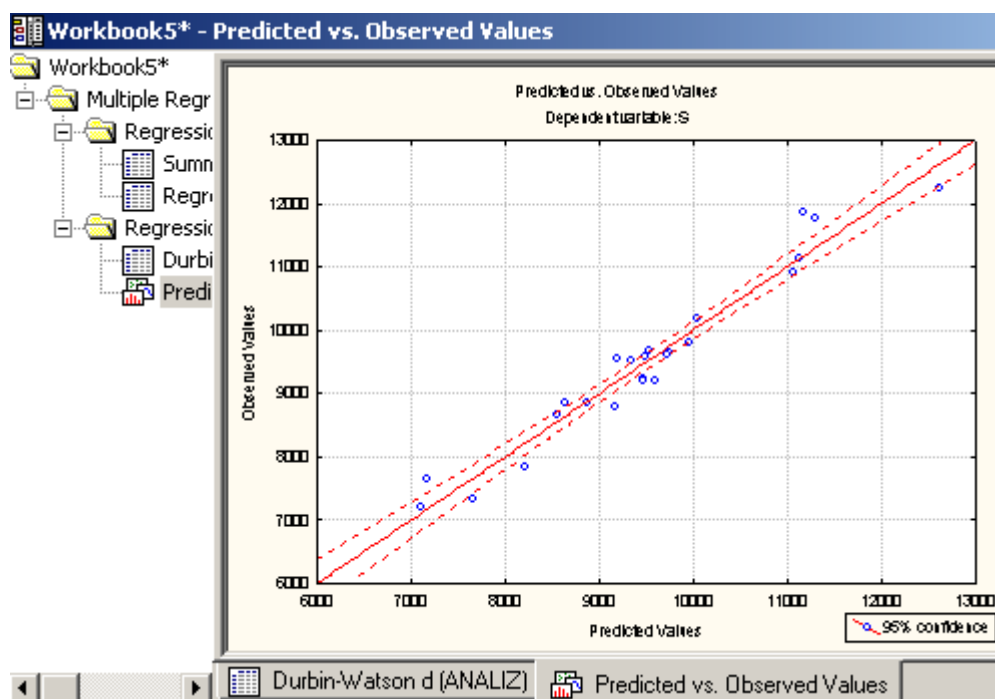
Workbook5* - Durbin-Watson d (ANALIZ)

- Workbook5*
- └─ Multiple Regr
- └─ Regression
- └─ Summ
- └─ Regr
- └─ Regression
- └─ Durbi

Durbin-Watson d (ANALIZ) and serial correlation of residuals			
	Durbin-Watson d	Serial Corr.	
Estimate	1,672399	0,141145	

Приложение 14

Наблюдаемые и прогнозируемые значения



СОДЕРЖАНИЕ

ВВЕДЕНИЕ.....	3
СПРАВОЧНЫЕ ДАННЫЕ.....	4
1. КРАТКИЕ СВЕДЕНИЯ ИЗ ТЕОРИИ СТАТИСТИКИ	4
2. БИРЖА	7
3. ПОЛЬЗОВАТЕЛЬСКИЙ ФОРМАТ В EXCEL	8
Лабораторная работа №1. ПОСТРОЕНИЕ ГРАФИКОВ.....	10
1. СОЗДАНИЕ ГРАФИКОВ	10
2. ФОРМАТИРОВАНИЕ ГРАФИКОВ	17
3. ГРАФИК БИРЖЕВЫХ ЦЕН.....	22
Лабораторная работа №2. РЕГРЕССИОННЫЙ АНАЛИЗ.....	27
1. МНОЖЕСТВЕННАЯ ЛИНЕЙНАЯ РЕГРЕССИЯ	27
2. ИНДИВИДУАЛЬНОЕ ЗАДАНИЕ	31
СПИСОК ЛИТЕРАТУРЫ	32
ПРИЛОЖЕНИЯ	33